Tabular Data

In machine learning, the most common type of data you'll encounter is **tabular data**—that is, data that is arranged in a data *table*. This is essentially the same format as you work with when you look at data in a spreadsheet.

Here's an example of tabular data showing some different clothing products and their properties:

SKU	Make	Color	Quantity	Price
908721	Guess	Blue	789	45.33
456552	Tillys	Red	244	22.91
789921	A&F	Green	387	25.92
872266	Guess	Blue	154	17.56

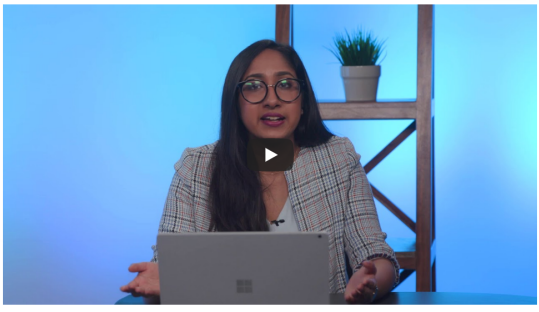
Notice how tabular data is arranged in **rows** and **columns**.

QUESTION 1 OF 2

Looking at the table above, can you figure out what the **rows** vs. **columns** are for?

- ✓ Each row describes a single product (e.g., a shirt), while each column describes a property the products can have (e.g., the color of the product)
- Each column describes a single product (e.g., a shirt), while each row describes a property the products can have (e.g., the color of the product)

SUBMIT



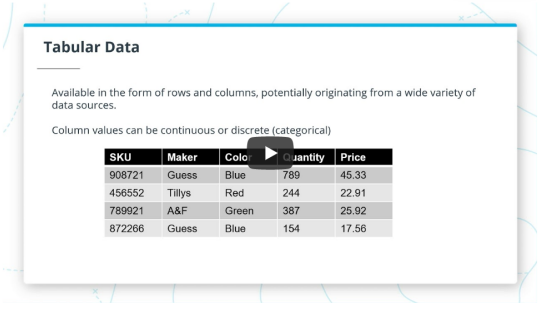
QUESTION 2 OF 2

Below are the components of a table. What does each of these components represent?

Submit to check your answer choices!

WHAT IT REPRESENTS	COMPONENT
An item or entity.	Row
A property that the items or entities in the table can have.	Column
A single value.	Cell

SUBMIT



Vectors

It is important to know that in machine learning we ultimately always work with numbers or specifically *vectors*.

> *A **vector** is simply an array of numbers, such as `(1, 2, 3)`—or a nested array that contains other arrays of numbers, such as `(1, 2, (1, 2, 3))`.*

Vectors are used heavily in machine learning. If you have taken a basic course in linear algebra, then you are probably in good shape to begin learning about how they are used in machine learning. But if linear algebra and vectors are totally new to you, there are some great free resources available to help you learn. You may want to have a look at Khan Academy's excellent introduction to the topic here or check out Udacity's free Linear Algebra Refresher Course.

For now, the main points you need to be aware of are that:

- All non-numerical data types (such as images, text, and categories) must eventually be represented as numbers
- In machine learning, the numerical representation will be in the form of an *array of numbers*—that is, a *vector*

As we go through this course, we'll look at some different ways to take non-numerical data and *vectorize* it (that is, transform it into vector form).

NEXT